

# Identifying “Useful” Fitness Models: Balancing the Benefits of Added Complexity with Realistic Data Requirements in Models of Individual Plant Fitness

Trace E. Martyn,<sup>1,\*</sup> Daniel B. Stouffer,<sup>2</sup> Oscar Godoy,<sup>3</sup> Ignasi Bartomeus,<sup>4</sup> Abigail I. Pastore,<sup>1</sup> and Margaret M. Mayfield<sup>1</sup>

1. School of Biological Sciences, University of Queensland, Brisbane, Queensland 4072, Australia; 2. Centre for Integrative Ecology, University of Canterbury, School of Biological Sciences, Christchurch, Canterbury 8041, New Zealand; 3. Departamento de Biología, Instituto de Universitario de Investigación Marina (INMAR), Universidad de Cádiz, 11510 Puerto Real, Spain; 4. Departamento Ecología Integrativa, Estación Biológica de Doñana (EDB-CSIC), Avda. Américo Vespucio 26, Isla de la Cartuja 41092, Sevilla, Spain

Submitted September 1, 2020; Accepted December 3, 2020; Electronically published March 4, 2021

Online enhancements: appendixes. Dryad data: <https://doi.org/10.5061/dryad.zs7h44j7f>.

**ABSTRACT:** Direct species interactions are commonly included in individual fitness models used for coexistence and local diversity modeling. Though widely considered important for such models, direct interactions alone are often insufficient for accurately predicting fitness, coexistence, or diversity outcomes. Incorporating higher-order interactions (HOIs) can lead to more accurate individual fitness models but also adds many model terms, which can quickly result in model overfitting. We explore approaches for balancing the trade-off between tractability and model accuracy that occurs when HOIs are added to individual fitness models. To do this, we compare models parameterized with data from annual plant communities in Australia and Spain, varying in the extent of information included about the focal and neighbor species. The best-performing models for both data sets were those that grouped neighbors based on origin status and life form, a grouping approach that reduced the number of model parameters substantially while retaining important ecological information about direct interactions and HOIs. Results suggest that the specific identity of focal or neighbor species is not necessary for building well-performing fitness models that include HOIs. In fact, grouping neighbors by even basic functional information seems sufficient to maximize model accuracy, an important outcome for the practical use of HOI-inclusive fitness models.

**Keywords:** species identity, higher-order interactions, functional traits, competition, annual plants, nonlinear responses.

## Introduction

Natural communities are incredibly complex. In fact, the more we study them, the clearer it becomes that we have much to learn about how they function and persist given their diversity and complexity. As such, our ability to accurately predict patterns of diversity and composition at local community scales is still fairly limited (Kimball et al. 2016; Houlahan et al. 2017; Maris et al. 2018). Such predictive capacity is important, however, for ensuring accurate understanding of community ecology and applying individual fitness models (i.e., models of lifetime fecundity) to a range of conservation and restoration problems.

Understanding how species interactions influence individual fitness and, hence, population dynamics is central to most models of community-level diversity (Volterra 1928; Chesson 2000a). Many of these models, however, are notoriously poor predictors of coexistence in highly diverse communities (Wootton 1994; White et al. 2006; Engel and Weltzin 2008). A reason for this is that such models often employ strong simplifying assumptions. One common assumption is that the only type of species interaction important for individual fitness and coexistence outcomes is direct competition between species pairs. There is increasing evidence, however, that other types of interactions are also important for structuring community diversity, including facilitation (Martorell and Freckleton 2014; Bulteri et al. 2016; Bimler et al. 2018), indirect competitive effects (Soliveres et al. 2015; Gallien et al. 2017; Godoy et al. 2017; Matías et al. 2018), and higher-order interactions (Billick and Case 1994; Bairey et al. 2016; Mayfield and

\* Corresponding author. Present address: School of Natural Resources and the Environment, University of Arizona, Tucson, Arizona 85721; email: [martyn.ecology@gmail.com](mailto:martyn.ecology@gmail.com).

**ORCID:** Martyn, <https://orcid.org/0000-0002-3015-484X>; Stouffer, <https://orcid.org/0000-0001-9436-9674>; Godoy, <https://orcid.org/0000-0003-4988-6626>; Bartomeus, <https://orcid.org/0000-0001-7893-4389>; Pastore, <https://orcid.org/0000-0002-2656-1725>; Mayfield, <https://orcid.org/0000-0002-5101-6542>.

Am. Nat. 2021. Vol. 197, pp. 415–433. © 2021 by The University of Chicago. 0003-0147/2021/19704-60167\$15.00. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which permits non-commercial reuse of the work with attribution. For commercial use, contact [journalpermissions@press.uchicago.edu](mailto:journalpermissions@press.uchicago.edu). DOI: 10.1086/713082

Stouffer 2017*b*). Further, a recent paper by Clark and colleagues (2019) found that models of intermediate complexity are the best at predicting local patterns of diversity. Their careful comparison of models containing different amounts of information made an important point: there is often a trade-off between how much biological information we include in models and their accuracy. It remains unclear, however, which biological details (rather than just how many) are the most important for accurately modeling and predicting coexistence outcomes and patterns of diversity.

In recent years, there has been a rapid increase in the number of studies exploring the importance of a wide range of species interactions to coexistence and diversity modeling (Goldberg and Werner 1983; Chesson 2000*b*; Michalet et al. 2014; Adler et al. 2018). One group of interactions, higher-order interactions (HOIs), has become of particular interest, given increasing evidence that they are strong and important for accurately modeling fitness and population growth rates in at least some plant communities (Mayfield and Stouffer 2017*b*). HOIs, defined here in the context of individual fitness models, are the nonadditive, cumulative effects of interactions between neighboring individuals on a focal individual's fitness (Billick and Case 1994; Wootton 1994). HOIs emerged as a way to formalize what has been lumped under the term "diffuse competition" (Mack and Harper 1977; Fowler 1981; Moen 1989). Diffuse competition (or "interaction milieu," as it is called by McGill et al. [2006]) is defined by MacArthur (1972) as competition by a constellation of species. The effects of the constellation of species are more specifically defined when using the concept of HOIs as the specific effect of the presence or density of a third species on the interaction between a species pair. Further, by this definition, HOI effects are specific to particular combinations of individuals of different species (Vandermeer and May 1969; Wootton 1994; Levine et al. 2017; Mayfield and Stouffer 2017*b*).

HOIs have been experimentally demonstrated on protozoa (Vandermeer and May 1969; Case and Bender 1981) and in food web analysis using experimental ponds (Morin et al. 1988; Wissinger and Mcgrady 1993; Paterson et al. 2015). Nearly all research on HOIs in ecology, however, has been based on theoretical, simulated data (Bailey et al. 2016; Levine et al. 2017) or has used experimental microcosms (Wilbur 1972; Neill 1974; Billick and Case 1994). To our knowledge, Mayfield and Stouffer (2017*b*) are the first to statistically evaluate the importance of HOIs in natural plant communities composed of many species, though other studies have now shown their importance in other natural communities as well (Li et al. 2020). Though including HOIs improved model-based predictions across these studies by adding flexibility to

the models, one potential downside is that they do not directly improve our understanding of the drivers of complex interactions. Indeed, the ecological mechanisms behind HOIs in plant communities are largely unknown and likely system specific (Kleinhesselink et al. 2020). Simulations by Letten and Stouffer (2019), however, indicate that HOIs routinely emerge from consumer-resource dynamics, indicating that shared use of resources (e.g., light or nutrients) may often be involved. Despite the lack of direct mechanistic insights, HOIs can provide useful insights about the complexity of species interactions and the role of diversity in competition. For this reason, it is valuable to improve our understanding of how to most effectively incorporate them into the individual fitness models that are foundational to the study of population dynamics and coexistence.

The predominant approach to inferring HOIs is statistically problematic in that it incorporates information about all pairs of neighboring species individually, resulting in a very large number of model terms. By adding such extensive detail, data requirements become massive for the study of HOIs in diverse systems. In fact, the data demands for models that incorporate even a fraction of all potential HOIs is so large, the tractability of this approach—or the inclusion of HOIs in individual fitness modeling of natural communities—has been questioned (Pomerantz 1981; Abrams 1983; Levine et al. 2017; Al-Adwani and Saavedra 2019). As the number of species in a community increases, the data requirements for parameterizing a full HOI-inclusive model (including data on all possible HOIs) grow exponentially and become impractical to collect. We note, however, the common misconception that this approach can be used only if there are data available for all species combinations—which is not true. In most cases, HOI-inclusive individual fitness models include a small fraction of potential HOIs (Mayfield and Stouffer 2017*b*; Li et al. 2020) because many combinations of species do not occur in natural communities and because most field data sets capture combinations of common species more completely than uncommon combinations. Though most efforts to parameterize HOI-inclusive models include only a fraction of possible HOI terms, no studies have tested whether particular types of HOIs such as intra- or interspecific HOIs are more important to include than others for model accuracy (Mayfield and Stouffer 2017*b*). One way to make the value of this added ecological complexity more tractable is to identify which details about interaction neighborhoods are actually needed to merit the benefits of HOI-inclusive models. For example, is it important to include individual model terms for each set of unique neighbors? Or can we categorize neighbors in ecologically sensible ways without losing the value added from HOIs overall? By identifying

which aspects of interaction neighborhoods are important and which are not, we can take an evidence-based approach to reducing the number of model terms and data requirements while maintaining a high level of biological realism.

A common practice in the study of drivers of biodiversity maintenance is to track the identities of focal individual species and neighbors to estimate the effect of species' interactions on individual fitness outcomes (Chesson 2000b; Adler et al. 2007). Mathematically and statistically, however, there is no requirement that species interactions differ based on species identity because no information on species differences (e.g., vital rates) are involved in calculating individual fitness (Hubbell 2001; Adler et al. 2007). Given this, there are many ways we could group species in HOI-inclusive models. Many such approaches have been explored for models based on direct interactions alone (Weigelt et al. 2002; Straub and Snyder 2006; Eskelinen 2008), with particular focus on tree communities (Callaway et al. 2003; Uriarte et al. 2004; Kaitaniemi and Lintunen 2010; L  bbe et al. 2015). For instance, Uriarte et al. (2004) found that for more than half of the focal species, grouping neighbor species was the best-fitting model, indicating that all neighboring plants acted similarly on the focal individual. Though the importance of neighbor identity has been well studied for fitness models that account for direct interactions, none, to our knowledge, has explored the same for fitness models that explicitly include HOIs. As past studies have found that direct interactions and HOIs do not capture the same variance (Mayfield and Stouffer 2017b), it is valuable to consider the importance of neighbor identity for HOI-inclusive models rather than assuming that identity will have the same relationship with HOIs as direct interactions alone. Following results from Uriarte et al. (2004) for tree communities, for instance, we start with the simplest option—a diffuse competition model. In this model, we assume that all neighboring plants are equivalent with respect to their competitive effect and, therefore, that their abundance can be summed into a single expression of neighbor abundance. Here, we also explore models that allow increasing amounts of complexity, and we group individual neighbors in ecologically sensible ways, including by rarity, functional role in the community, or origin status, or on the basis of taxonomic relationships (McGill et al. 2006).

Grouping rare species together is a comparatively quick way to increase model tractability by reducing the number of free parameters requiring statistical estimation in the model while limiting the amount of ecological information lost. Rare species often establish weak interactions within competitive networks that appear to stabilize community dynamics overall (McCann et al. 1998; V  zquez et al. 2007; Gellner and McCann 2016). Locally rare spe-

cies, however, may contribute so few individuals to a community that the species-specific impact on the population growth rate of a neighboring species is negligible. We can, therefore, test whether rare species can be grouped in models of individual fitness or, indeed, removed as competitors from the models altogether without losing the benefits of including HOIs.

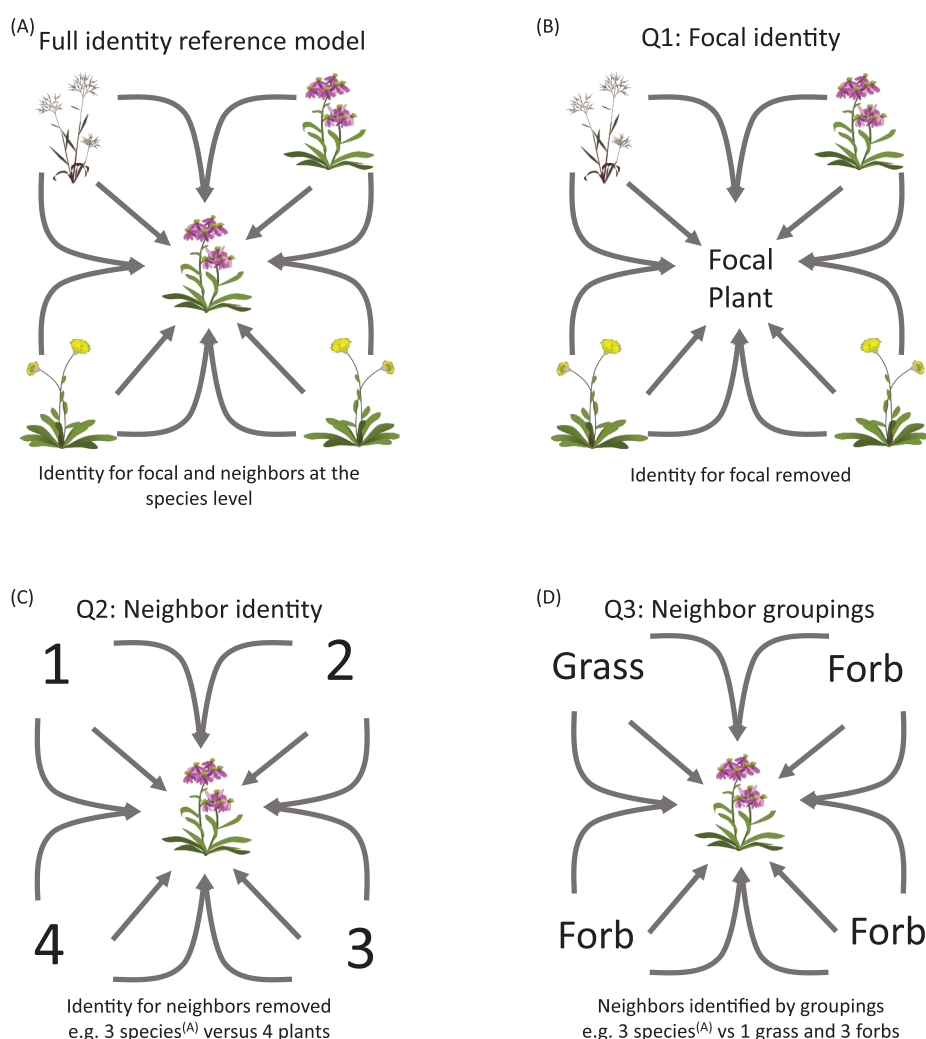
Using broad categories such as functional group or origin status (native/exotic) provides a way to group individuals that may play similar roles within the community but limit the amount of detail needed to explain general patterns. For example, in herbaceous communities, categorizing species by life form (i.e., grass or forb) can capture important differences in life-history strategy, traits (see Craine et al. 2001; Tjoelker et al. 2005; Ravenek et al. 2016), responses to nutrient addition (see Li et al. 2016; You et al. 2017), and water availability (P  rez-Ramos et al. 2019). Further, we may expect exotic and native species to have different functional traits (Ordonez et al. 2010) and differing competitive abilities (Levine et al. 2003; Godoy et al. 2014). For example, Ordonez et al. (2010) found that exotic species tend to have small seeds (which could influence wider dispersal) and large specific leaf area (SLA; which could improve ability to compete for light). Combining origin status with life form to create narrower categories (e.g., a native, annual grass vs. an exotic, perennial forb) thus seems an ecologically sensible way to create finer ecological groups that provide important information about how individuals interact without dramatically increasing model complexity. It may also prove beneficial to add a bit more detail by including specific functional trait data in species groupings (D  az et al. 2016; Levine 2016). Alternatively, grouping species by taxonomic relatedness, such as family, offers a potentially simple way to capture key species differences as it is a measure of evolutionary history that potentially predicts trait similarity between species in communities (Qian and Jiang 2014; also reviewed in Webb et al. 2002; Cavender-Bares et al. 2009; but see Wilcox et al. 2018).

In this paper, we explore a range of approaches to simplifying annual plant individual fitness models that include more biological information than direct competition alone, focusing on models that include higher-order interactions and facilitation as well as direct competition. Specifically, we ask, the following: (1) Is considering the identity of focal species important for accurately modeling individual fitness (i.e., lifetime fecundity) in annual plant communities? (2) How important is considering the identity of neighborhood members for the accuracy of individual fitness models? Is it worthwhile to include details about interactions with rare species, or can they be grouped or ignored entirely without losing model accuracy? And, (3) Can we identify generalizable functional

and taxonomic approaches to grouping neighbor species that reduce model complexity while maintaining ecologically important variation associated with direct and higher-order interactions?

To answer these questions, we use the Mayfield and Stouffer (2017b) higher-order interaction model framework to model individual fecundity of focal plants. We split species identity in fitness models into two parts, the focal identity (question 1) and the neighbor identity (questions 2 and 3; fig. 1). For question 1, we test the importance of focal species identity by iteratively including a

focal identity term in models that include combinations of direct and higher-order interactions (fig. 1B; table 1). To ensure the generalizability of our findings, we then use a variety of approaches to assess importance. First, we define importance as an increase in model accuracy (lower value) based on comparing root mean square error (RMSE) for model fit to testing and training data sets ( $RMSE_{train}$  and  $RMSE_{test}$ ). By comparing RMSE across different model specifications, researchers can balance how much accuracy they are willing to compromise to simplify the models. In addition, we also provide two different but commonly



**Figure 1:** Conceptualizations of identity analyses showing a simple neighborhood with a focal individual (center pink flower) and four neighboring plants. *A*, The standard neighborhood is where the focal plant and the neighbors are identified by their species; this is the neighborhood that all other analyses will be compared against. *B*, To analyze the importance of the identity of the focal species, the standard neighborhood is compared to a neighborhood where the focal species identity is removed (here indicated as “focal plant”). *C*, The standard neighborhood is compared to a neighborhood where the neighbor species identity is removed and neighbors are accounted for as just counts of plants (abundance, represented by the numbers 1–4). *D*, The standard neighborhood is compared to a neighborhood where the neighboring plants are grouped by taxonomic or functional identity (shown is an example identifying differences in life form, i.e., grass vs. forb).

Table 1: Focal analysis table

Model	Mathematical formula	Statistical interpretation	Biological interpretation
No ID (direct)	$F_{m_X} \{N\} = \lambda_X e^{D_{m_X} \{N\}} ,$ $D_{m_X} \{N\} = -\sum_{j=1}^s \alpha_{Xj} N_j$	Intercept and direct interaction slopes do not vary with focal identity	Focal species do not differ in intrinsic fecundity or their direct responses to neighbor abundance
No ID (direct and HOI)	$F_{m_X} \{N\} = \lambda_X e^{D_{m_X} \{N\} + (H_{m_X} \{N\})} ,$ $D_{m_X} \{N\} = -\sum_{j=1}^s \alpha_{Xj} N_j ,$ $H_{m_X} \{N\} = -\sum_{j=1}^s \beta_{Xjj} \frac{N_j(N_j - 1)}{2} - \sum_{j=1}^s \sum_{k=j+1}^s \beta_{Xjk} N_j N_k$	Intercept and direct and higher-order interaction slopes do not vary with focal identity	Focal species do not differ in intrinsic fecundity or their direct and higher-order responses to neighbor abundance
Impartial $\alpha$ (direct)	$F_{m_i} \{N\} = \lambda_i e^{D_{m_i} \{N\}} ,$ $D_{m_X} \{N\} = -\sum_{j=1}^s \alpha_{Xj} N_j$	Intercept and direct interaction slopes do vary with focal identity, but higher-order interaction slopes do not	Focal species do differ in intrinsic fecundity but do not differ in their direct response to neighbor abundance
Impartial $\alpha$ and $\beta$ (direct and HOI)	$F_{m_i} \{N\} = \lambda_i e^{D_{m_i} \{N\} + (H_{m_i} \{N\})} ,$ $D_{m_X} \{N\} = -\sum_{j=1}^s \alpha_{Xj} N_j ,$ $H_{m_X} \{N\} = -\sum_{j=1}^s \beta_{Xjj} \frac{N_j(N_j - 1)}{2} - \sum_{j=1}^s \sum_{k=j+1}^s \beta_{Xjk} N_j N_k$	Intercept does vary with focal identity, but direct and higher-order interaction slopes do not	Focal species do differ in intrinsic fecundity but do not in their direct and higher-order responses to neighbor abundance
Impartial $\beta$ (direct and HOI)	$F_{m_i} \{N\} = \lambda_i e^{D_{m_i} \{N\} + (H_{m_i} \{N\})} ,$ $D_{m_i} \{N\} = -\sum_{j=1}^s \alpha_{ij} N_j ,$ $H_{m_X} \{N\} = -\sum_{j=1}^s \beta_{Xjj} \frac{N_j(N_j - 1)}{2} - \sum_{j=1}^s \sum_{k=j+1}^s \beta_{Xjk} N_j N_k$	Intercept and direct interaction slopes do vary with focal identity but do not for higher-order interaction slopes	Focal species do differ in intrinsic fecundity and their direct response to neighbor abundance but do not in their higher-order response to neighbors
Full model (direct)	$F_{m_i} \{N\} = \lambda_i e^{D_{m_i} \{N\}} ,$ $D_{m_i} \{N\} = -\sum_{j=1}^s \alpha_{ij} N_j$	Intercept and direct interactions slopes do vary with focal identity	Focal species do differ in intrinsic fecundity and their direct response to neighbor abundance
Full model (direct and HOI)	$F_{m_i} \{N\} = \lambda_i e^{D_{m_i} \{N\} + (H_{m_i} \{N\})} ,$ $D_{m_i} \{N\} = -\sum_{j=1}^s \alpha_{ij} N_j ,$ $H_{m_i} \{N\} = -\sum_{j=1}^s \beta_{ijj} \frac{N_j(N_j - 1)}{2} - \sum_{j=1}^s \sum_{k=j+1}^s \beta_{ijk} N_j N_k$	Intercept and slopes for direct and higher-order interactions do vary with focal identity	Focal species do differ in intrinsic fecundity and their direct and higher-order response to neighbor abundance

Note: Shown are the names of the models in the focal-species identity analysis, how those models are modified (represented by mathematical equation), the statistical interpretation, and the biological interpretation. In the analysis, we added in focal identity into the models for the intrinsic fecundity ( $\lambda$ ), direct interaction ( $\alpha$ ), and higher-order interaction (HOI;  $\beta$ ) terms iteratively. When focal identity is present in the term, the term has a subscript  $i$ ; if focal identity is absent, the subscript is  $X$ .



used metrics for model fit that penalize for number of parameters: Akaike's information criterion (AIC) and Bayesian information criterion (BIC). These model selection heuristics objectively select the best model with different criteria. Overall, the models that result in the best fit inform if and for which interactions focal identity is important. For questions 2 and 3, we compare models with different neighbor identities (e.g., plant abundance only [fig. 1C] and neighboring species groupings based on rarity or functional identity, such as life form [fig. 1D]). If we find an improvement in model performance with certain groupings, this provides insight into ways of grouping species that improve model tractability while retaining key biological information about neighboring species. We ran the above analyses with data from two distinct annual plant communities—one from Australia and the other from Spain—in order to determine not only which simplification approach is best but also which is most generalizable across contrasted annual plant systems.

## Methods

### *Data Sets*

We ran all analyses on two independent data sets collected from two distinct study systems, one in Australia and one in Spain. Both data sets come from annual plant-dominated communities growing naturally in Mediterranean climate regions. We describe each study system and data-collection protocols below.

**Australia.** Our first data set was collected in 2013 in two Western Australia nature reserves, Kunjin (lat. 32°21'19.31"S, long. 117°45'42.3"E) and Bendering (lat. 32°23'06.1"S, long. 118°22'42.4"E). Both reserves are dominated by York gum-jam woodlands, which support an understory of mixed native and exotic annual grasses and forbs (Dwyer et al. 2015). Across both reserves, we tracked seed production of individual plants from six focal species. Three species were found at both reserves, whereas the others were found in only one reserve or the other. In each reserve, 30 × 30-cm plots were marked out around arbitrarily selected individuals of each focal species. There were 35 focal individuals for each focal species at each reserve in which it was found, for a total of 945 focal individuals across all six species and both reserves. Five of the individuals selected from each species were randomly assigned to an intrinsic fecundity treatment, in which all neighbors were weeded in a 7.5-cm-radius circle around the focal plant. The intrinsic fecundity treatment served to quantify the number of seeds for an individual that was receiving no competition (i.e., a baseline fecundity). The remaining 30 plots were competition plots, in which neighborhoods were thinned to contain cer-

tain combinations of neighbors. Distinct neighborhood-thinning treatments were not distinct enough to be different factors and so are not considered here. For more details on neighborhood thinning, see Mayfield and Stouffer (2017b). In a 7.5-cm radius around each focal individual, we recorded the identity and abundance of all neighboring plants left after thinning. Neighborhoods around focal plants included 0–31 individuals from zero to eight different species. Across all plots, there were 45 neighboring species represented, including the six focal species (see table S1 for the list of species with sample sizes; tables S1–S5 are available in online apps. S1–S4).

At the end of the growing season, we counted all flowers (inflorescences for Asteraceae species) on each focal plant and collected seeds from one to three flowers (inflorescences) per plant. To estimate total seed production per plant, we then averaged the number of seeds produced per flower and multiplied this number by the number of flower heads on each individual. Herbivory, plant death, or seed release prior to collection prevented the use of 172 focal plants, reducing our final data set to 773 focal individuals from six focal species across two reserves. See Mayfield and Stouffer (2017b) for further field methods. Fecundity and neighborhood data are from Mayfield and Stouffer (2017a).

**Spain.** Our second data set was collected from an annual grassland system in Caracoles Estate within Doñana National Park in southwest Spain (lat. 37°04'01.0"N, long. 6°19'16.2"W). In September 2015, we established nine 8.5 × 8.5-m plots over a humidity and salinity gradient within a 1 km × 200 m area. Plots were blocked into three locations along a topographical gradient (upper, middle, and lower) with, on average, 300 m between each block and 15 m between each plot (fig. S1; figs. S1–S12 are available in online apps. S1–S4). Each plot was subdivided into 36, 1 × 1-m subplots, and we measured one focal individual per species within each subplot for a maximum of 324 focal individuals per species. Surrounding each focal individual, we counted and identified the neighbor species within a 7.5-cm-radius area. During senescence, we recorded seed production for each individual that germinated by counting the number of fruits on each individual and multiplying that value by the average number of seeds per fruit for each species at the site. In total, we collected fecundity and neighborhood data for 1,751 focal individuals from 17 focal species at this site. For this analysis, we removed focal species with less than 20 occurrences, resulting in a data set with 1,694 individuals from the 12 most common species with 17 potential neighbors (see table S1 for the list of focal and neighboring species with sample sizes). See Lanuza et al. (2018) and García-Callejas et al. (2020) for more in-depth field methods.

### Individual Fitness Model

We quantified the effects of neighbors on the fecundities of focal individuals using the negative binomial individual fitness model framework presented by Mayfield and Stouffer (2017b):

$$F_{m_i}|\{N\} = \lambda_i e^{(D_{m_i}|\{N\}) + (H_{m_i}|\{N\})}, \quad (1)$$

where  $F_{m_i}$  is the fecundity (measured as the number of seeds a plant produced or was estimated to produce) of each focal individual  $m$  of the focal species  $i$  in the presence of a set of neighboring individuals  $\{N\}$ . Intrinsic fecundity (i.e., fecundity in the presence of no neighbors) of focal species  $i$  is represented above by  $\lambda_i$ . We used  $D_{m_i}$  to partition the direct effects of neighbors on the focal individual from the higher-order effects  $H_{m_i}$ . Direct effects are modeled more specifically as

$$D_{m_i}|\{N\} = - \sum_{j=1}^s \alpha_{ij} N_j, \quad (2)$$

where the direct effects ( $\alpha_{ij}$ ) capture the per capita effect of the density  $N_j$  of neighbors of species  $j$  on the focal individual  $m_i$ . Higher-order effects were modeled similarly:

$$H_{m_i}|\{N\} = - \sum_{j=1}^s \beta_{ijj} \frac{N_j(N_j - 1)}{2} - \sum_{j=1}^s \sum_{k=j+1}^s \beta_{ijk} N_j N_k, \quad (3)$$

where higher-order effects  $\beta_{ijk}$  are the collective effects of neighboring individuals of species  $j$  and/or species  $k$  on the focal individual, resulting in nonadditive effects of neighbor densities (e.g.,  $N_j N_k$ ) on the fecundity of the focal individual. For intraspecific higher-order effects, we used the term  $\beta_{iii}$  (which captures the quadratic effect of intraspecific density on the focal individual's fecundity), and interspecific HOIs could be either  $\beta_{ijj}$  or  $\beta_{ijk}$ .

### Analyses

For our analyses, we used a model comparison approach to determine how the identity of focal and neighbor species affects model accuracy. Here, we determine model accuracy using four metrics. We used  $k$ -fold cross-validation (Zhang 1993) to compare the fit of the model to the data and the ability of the model to predict withheld data. We split each data set into five stratified sections using the `createFolds()` function in the `caret` package in R (Kuhn 2008). We fit the model with four of the five folds and then predicted the remaining fold using the withheld data as a testing data set. To calculate the fit to the training data and the ability to predict the testing data, we used the RMSE, which is expressed in the same units used to express the linear predictor— $\log(\text{seeds})$  in our negative binomial model—and can

quantitatively inform us of how much error is added via model simplification. We calculated RMSE for the training data set as

$$\text{RMSE}_{\text{train}} = \sqrt{\text{mean}(\text{residuals}^2)}, \quad (4)$$

which is not penalized by the number of model terms, unlike AIC or BIC. Note that, for a set of nested generalized linear models, the one with more parameters will always show a smaller model deviance (i.e., the value being minimized during model fitting) and, hence, a larger log likelihood due to the relationship between deviance and log likelihood; however, the model with more parameters will not necessarily have a smaller RMSE due to heteroscedasticity or other variation in the data (McCullagh and Nelder 1989).

We also calculated  $\text{RMSE}_{\text{test}}$  based on the predicted values as

$$\text{RMSE}_{\text{test}} = \sqrt{\text{mean}((\text{observed values} - \text{predicted values})^2)}, \quad (5)$$

which we used as a measure of how well the model performs for withheld data. Using both  $\text{RMSE}_{\text{train}}$  and  $\text{RMSE}_{\text{test}}$ , we also estimated loss in predictive power of the model as

$$\text{proportion loss model predictive power} = 1 - \frac{\text{RMSE best fit model}}{\text{RMSE model to compare against}}, \quad (6)$$

so we could compare how much fit is lost by choosing a different model (perhaps with fewer parameters to estimate) compared to the best fit model (table S2). In the text, we present percentage loss, which is the proportion value multiplied by 100. We also used two model accuracy metrics that are penalized by the number of parameters in the model, AIC and BIC, with BIC having a stronger penalty for number of terms in the model than AIC (Brewer et al. 2016). For each analysis, we considered the best model to be the one that had the lowest AIC/BIC value by at least 2 points (Burnham and Anderson 2002). We present the average model accuracy metric values across all five folds of the data in the results section and in table 3.

To fit the models, we used the `manyglm()` function (package `mvabund`; Wang et al. 2012). Using this package, rank-deficient terms were dropped during model fitting. Models were fit separately for the Australian and Spanish data sets. In the Australian data set, because the sites were approximately 65 km away, for focal species sampled at different sites, a fixed effect of site was included in the models to account for differences between sites. A fixed effect for site was not used for the Spanish data set because the plots were relatively close in distance (<500 m). All analyses were

performed in R statistical software 3.5.1 (R Core Development Team 2018).

**Focal Identity Analysis.** To explore the importance of focal species identity to all terms in the model, we iteratively added the identity of the focal species to the  $\lambda$ ,  $\alpha$ , and  $\beta$  terms (intrinsic fecundity, direct interaction, and higher-order interactions, respectively), starting with the null expectation that focal identity is not important for model fit (“no ID (direct)” and “no ID (direct and HOI)” models in table 1). For those models that included only direct effects, we fixed all  $\beta$  terms to zero. We did not do a whole factorial design for this analysis in order to maintain biologically interpretable models. For example, we reason that it would be unrealistic to include focal species identity for HOI effects without also including focal species identity for the direct effects that are modified by those HOIs and the intrinsic fecundity for the focal individual. Therefore, our approach iteratively added complexity to each previous model, starting with focal identity excluded from all terms and then adding in focal species identity sequentially to the intrinsic fecundity term, direct interaction terms, and, finally, HOI terms.

**Neighbor Identity and Rarity Analysis.** To answer our second question about the importance of neighbor identity, we conducted a similar metric-based model comparison as above, comparing different neighbor groupings using the full direct-only model (“full model (direct)”) and the full HOI-inclusive model (“full model (direct and HOI)”), which included all species-level focal and neighbor identities. We first compared the full species identity model to the simplest model, abundance, in which neighbors were included just as total number of individuals around each focal individual (for the direct effects) and pair of individuals (for the higher-order effects). This simplified model allowed us to ask whether the abundances of

neighbors alone could capture the important information about a neighborhood on a focal individual’s fecundity when HOIs are included.

To explore the importance of neighbor species rarity, we calculated the total number of neighbors for each data set (3,582 for the Australian data set and 19,831 for the Spanish data set). We then grouped species that had abundances of less than 1% across the whole data set into a rare species grouping (28 of 45 neighbor species for the Australian data set and eight of the 17 neighbor species for the Spanish data set were classified as rare; starred in table S1). We then also ran the models with those rare species removed from the data set to see how removing the rare species would affect model fit. We compared the results for both the full direct-only model and the full HOI-inclusive model.

**Neighbor Grouping Analysis.** For our third question, exploring generalizable functional groupings, we compared the full higher-order inclusive model with all identities included with a selection of biologically motivated neighbor groupings. We grouped all species by life form, origin status, functional traits (see app. S2 for trait-clustering methods), and taxonomic relatedness (see app. S2 for methods; table 2; table S1). We did not test origin status for the Spanish data set because there was only one status (native). Overall, we tested seven distinct neighbor identity groupings against the models run with full species identity (groupings listed in table 2 along with the two rare groupings above).

## Results

### *Focal Identity*

The best models differed for our two data sets. For the Australian data set, the impartial  $\alpha$  (direct) model, which includes focal identity for only intrinsic fecundity ( $\lambda_i$ ),

**Table 2:** Groupings for neighbor identity analysis

Neighbor identity group	Definition	Example	Number of groups within data set	
			Australia	Spain
Life form	Grouped by whether species is graminoid or forb	Forb, grass	2	2
Origin status	Grouped by whether species is exotic or native to site	Exotic, native	2	1 <sup>a</sup>
Functional type	Grouped by life form, life history, and native status	Annual exotic grass, . . . , perennial native forb	4	2
Trait complex	Grouped by trait clustering of functional traits (see app. S2)	Cluster group 1, . . . , cluster group 4	7	2
Plant family	Grouped by plant family	Asteraceae, . . . , Goodeniaceae	18	9

Note: See table S1 for complete species list and assigned groupings.

<sup>a</sup> Because there was only one group for origin status in the Spain data set, this model type was not run for that data set.



was the best model based on having the lowest  $RMSE_{test}$ , AIC, and BIC values (table 3). The full model (direct and HOI) was the best-performing model based on only  $RMSE_{train}$ . Between the impartial  $\alpha$  (direct) model and the full model (direct and HOI), there was a 16% loss in predictive power based on  $RMSE_{train}$  and a 63% gain in predictive power based on  $RMSE_{test}$ . There was also a 12-fold increase in the number of terms for the model to estimate in the full model (direct and HOI) compared to the impartial  $\alpha$  (direct) model (694 and 55 terms, respectively; table 3).

For the Spanish data set, the best model based on  $RMSE_{train}$  and AIC was the full model (direct and HOI), within which focal identity was included for all model terms (table 3). When considering  $RMSE_{test}$  and BIC, however, the full model (direct) was the best-performing model. There was a 22% loss in predictive power between the full model (direct and HOI) and the full model (direct) based on  $RMSE_{train}$  and a 600% increase in predictive power based on  $RMSE_{test}$ . There were also 382 fewer parameters to estimate in full model (direct) compared to full model (direct and HOI).

#### Neighbor Identity and Rarity

For the Australian data set, the HOI-inclusive model that used the abundance grouping performed better in three of the four metrics (all but  $RMSE_{train}$ ) than did the HOI-inclusive species identity model (red and black points, respectively, in figs. 2 and S2). When comparing performance

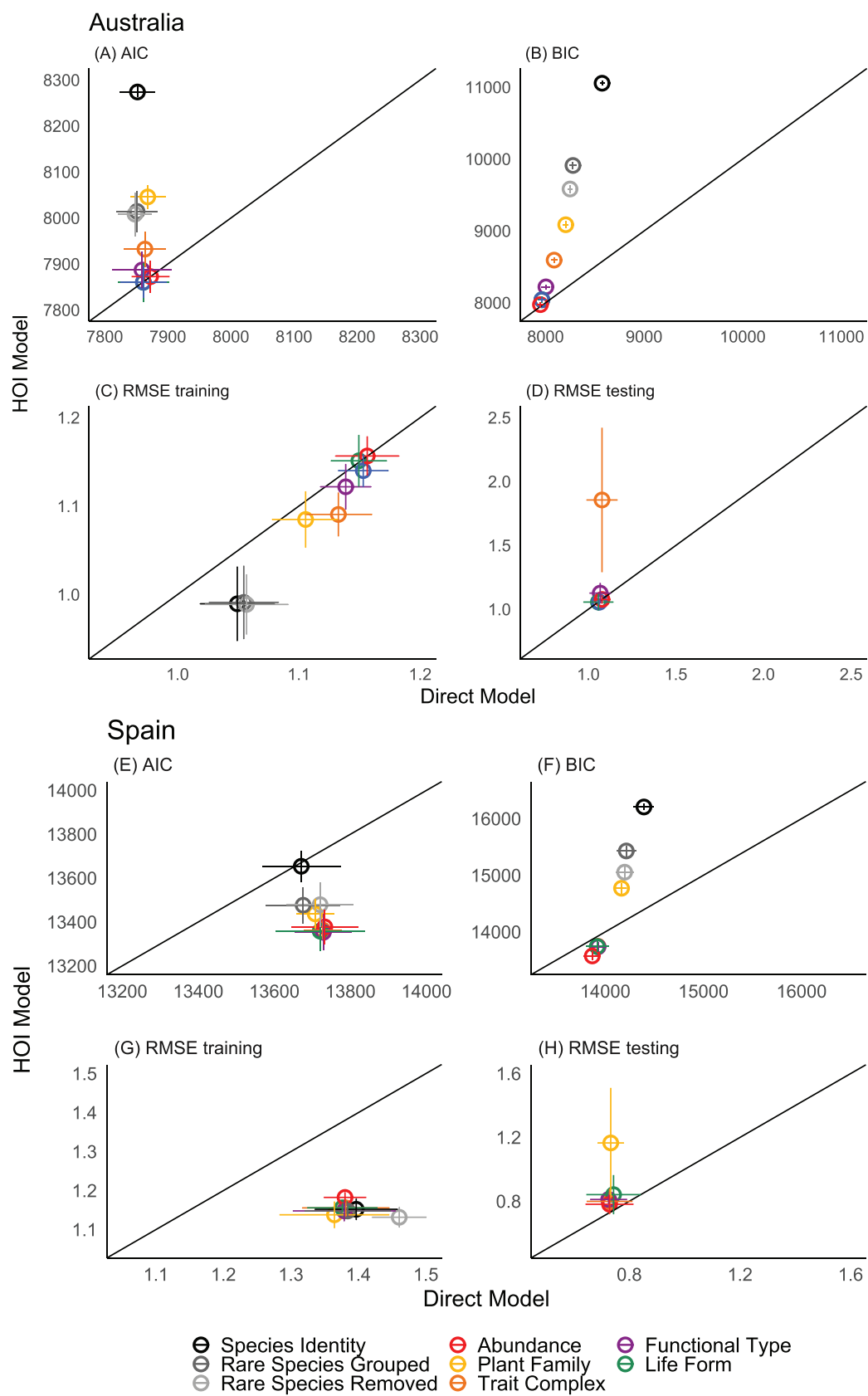
based on  $RMSE_{test}$ , the species identity model performed worse than the abundance models, with a 33% loss of predictive power for the direct-only model and a more than 20,000% loss for the HOI-inclusive model. Both the direct-only and HOI-inclusive abundance models had a 16% decrease in  $RMSE_{train}$  predictive power compared to the best-fitting HOI-inclusive species identity model; however, the abundance models had 97% fewer parameters to estimate compared to the HOI-inclusive species identity model (table S2).

Similar to the Australian data set, for the Spanish data set, the abundance grouping performed better than the species identity grouping for all accuracy metrics for the HOI-inclusive model except  $RMSE_{train}$  (red and black points, respectively, in figs. 2 and S3). For direct-only models, there was an improvement in accuracy based on BIC but lack of clear improvements in the other three metrics ( $RMSE_{train}$ ,  $RMSE_{test}$ , and AIC). When comparing  $RMSE_{train}$  between the abundance and species identity HOI models, there was a 3% loss in predictive power; however, the abundance model had 92% fewer parameters to estimate. Both the abundance and species identity direct-only models had a 20% loss in performance based on  $RMSE_{train}$  compared to the species identity HOI-inclusive model, though there were 94% and 71% fewer parameters to estimate in those models, respectively (table S2). There was more than a 300% decrease in predictive power and a 28% increase in number of parameters to estimate when comparing the  $RMSE_{test}$  between the direct-only and HOI-inclusive species identity models.

**Table 3:** Focal identity results

Data set, grouping	Coefficient	AIC	BIC	$RMSE_{train}$	$Loss_{train}$	$RMSE_{test}$	$Loss_{test}$
Australia:							
No ID (direct)	47	10,447	10,700	1.465	-.47	1.630	-.02
No ID (direct and HOI)	348	10,770	12,393	1.358	-.36	26.693	-24.61
Impartial $\alpha$ and $\beta$ (direct and HOI)	355	10,008	11,663	1.081	-.08	21.797	-19.91
Impartial $\alpha$ (direct)	55	<b>9,767</b>	<b>10,027</b>	1.155	-.16	<b>1.042</b>	<b>.0</b>
Impartial $\beta$ (direct and HOI)	459	10,069	12,208	1.046	-.02	12.352	-10.85
Full model (direct)	170	9,800	10,595	1.017	-.05	1.240	-.19
Full model (direct and HOI)	694	10,263	13,495	<b>.995</b>	<b>.0</b>	67.213	-63.50
Spain:							
No ID (direct)	18	19,812	19,915	1.279	-.11	1.892	-1.17
No ID (direct and HOI)	127	19,770	20,465	1.334	-.16	2.362	-1.72
Impartial $\alpha$ and $\beta$ (direct and HOI)	138	17,676	18,431	1.287	-.12	1.382	-.59
Impartial $\alpha$ (direct)	30	17,939	18,107	1.213	-.05	.967	-.11
Impartial $\beta$ (direct and HOI)	248	17,139	18,492	1.392	-.21	1.182	-.36
Full model (direct)	145	17,068	<b>17,862</b>	1.395	-.22	<b>.868</b>	<b>.0</b>
Full model (direct and HOI)	527	<b>16,996</b>	19,866	<b>1.146</b>	<b>.0</b>	6.023	-5.93

Note: Listed are the number of coefficients for the model as well as the Akaike information criterion (AIC), Bayesian information criterion (BIC), root mean square error (RMSE) for models fit to the training data set ( $RMSE_{train}$ ), RMSE for models fit to the testing data set ( $RMSE_{test}$ ), and proportion predictive power loss for those respective RMSE values ( $Loss_{train}$  and  $Loss_{test}$ ). Boldfaced values are those that are the best-performing models based on the respective accuracy metric. ID = focal identity; HOI = higher-order interactions.



For the rare groupings, the HOI-inclusive model with rare species removed improved predictive power based on  $RMSE_{train}$  by 0.4% over the HOI-inclusive models of species identity and rare species grouped and 6% over any direct-only models. Based on  $RMSE_{test}$ , the direct-only model with rare species grouped improved predictive power over rare species removed and species identity by 3% and 35%, respectively. When compared to the direct-only models using  $RMSE_{test}$ , the HOI-inclusive models performed poorly, with a greater than 2,000% loss in predictive power (table S2). The models with rare species grouped and rare species removed performed better than the species-identity model based on AIC and BIC, especially the HOI-inclusive model for the Australian data set (gray and black points, respectively, in figs. 2 and S2). By grouping and removing rare species, we reduce the number of parameters in the model by 31% and 43%, respectively, for the HOI-inclusive models and 41% and 45%, respectively, for the direct-only models when compared to the number of parameters in the full species identity HOI-inclusive and direct-only models (table S2).

Similar to the Australian data set, for the Spanish data set, based on  $RMSE_{train}$ , the HOI-inclusive model with rare species removed had a 1.4% and 0.7% improvement in predictive power over the HOI-inclusive models with species identity and rare species grouped, respectively, and a greater than 22% improvement in predictive power over any of the direct-only models (table S2). Using  $RMSE_{test}$ , the direct-only model with rare species removed was the best performing, improving in predictive power by 0.5% and 6.4% over the direct-only models with species identity and rare species grouped, respectively. The HOI-inclusive models had at least a 300% loss in predictive power based on  $RMSE_{test}$  compared to any of the direct-only models (table S2). Also similar to the Australian data set, the rare species categories for the Spanish data set performed better for AIC and BIC (fig. 2). Compared to the species-identity models, rare species removed and rare species grouped had 38% and 23% fewer parameters to estimate

in the HOI-inclusive models and 33% and 23% in the direct-only models.

### Neighbor Grouping

For the Australian data set, most neighbor groupings performed better than then full species-identity model based on  $RMSE_{test}$ , AIC, and BIC but not  $RMSE_{train}$  for the HOI-inclusive models (vertical spread of colored vs. gray/black points in fig. 2). For the direct-only models, there was improvement in model performance over species identity based on BIC for all neighbor groupings. However, improvement was less apparent based on  $RMSE_{train}$ ,  $RMSE_{test}$ , and BIC (horizontal spread of points in fig. 2). Compared to the species identity HOI-inclusive model, plant family and trait complex had a 9%–11% decrease in predictive power based on  $RMSE_{train}$  (for both HOI-inclusive and direct-only models), whereas functional type, life form, and origin status had a 13%–16% decrease for both direct-only and HOI-inclusive models. Compared to the full, HOI-inclusive species identity model, there were 63%, 76%, 88%, 94%, and 94% fewer parameters to estimate for the plant family, trait complex, functional type, life form, and origin status groupings, respectively (app. S1; table S2). For the direct-only models, compared again to the species identity HOI-inclusive model, there were more than 88% fewer parameters for all of the above groupings.

For the Spain data set, groupings appeared to perform as well as or better than the species identity model for  $RMSE_{train}$ ,  $RMSE_{test}$ , AIC, and BIC in the HOI-inclusive models (vertical spread in fig. 2). Similar to the Australian data set, for the direct-only models, the improvement in BIC is clear; however, for the other metrics there was little improvement in model performance (horizontal spread in fig. 2). The best-performing model based on  $RMSE_{train}$  was the plant family HOI-inclusive model. Compared to this best-performing  $RMSE_{train}$  model, the other groupings’ HOI-inclusive models (as well as the species identity model) had up to 2% loss in predictive power. The

**Figure 2:** Model accuracy metrics for neighbor groupings comparing direct-only models and models including higher-order interactions (HOI) for the Australian data set (A–D) and the Spanish data set (E–H). Shown are the mean and standard error ( $\pm 2$ ) values for four different model metrics: Akaike information criterion (AIC; A and E), Bayesian information criterion (BIC; B and F), root mean square error (RMSE) for the model fit to the training data set in the  $k$ -folds cross-validation analysis ( $RMSE_{train}$ ; C and G), and RMSE for the model predicting to withheld data in the  $k$ -folds cross-validation analysis ( $RMSE_{test}$ ; D and H). Metrics for the direct-only models are on the X-axis and values for the HOI-inclusive model on the Y-axis. Points above the 1:1 line indicate that the direct-only model fit better (lower value) for that metric, while those below the line indicate that the HOI-inclusive model fit better. Different colors represent the different neighborhood groupings used in the analyses: full species identity of the neighbors (species identity), grouping rare species together into a rare group (rare species grouped), removing the rare species (rare species removed), considering neighbors only by abundance (abundance), and the groupings list in table 2. For the Australian data set, for  $RMSE_{test}$  (D), the species identity, rare species grouped, rare species removed, and plant family grouping HOI-inclusive model values are much greater than those shown and are not included in this figure to improve clarity for the other groupings (see fig. S2 for the full image with the above groupings included). Similarly for the Spanish data set, for  $RMSE_{test}$  (H), the species identity, rare species grouped, and rare species removed grouping HOI-inclusive model values are much greater than those shown and are not included (see fig. S3 for the full image).

direct-only models for all groupings had a 20%–23% loss in predictive power compared to the best-performing  $\text{RMSE}_{\text{train}}$  model. Based on  $\text{RMSE}_{\text{test}}$ , the best-performing model was the species identity direct-only model. The direct-only models for all the groupings had up to a 1% loss in predictive power based on  $\text{RMSE}_{\text{test}}$  compared to the species identity direct-only model, while the HOI-inclusive models had a 61%, 10%, 12%, and 16% loss in predictive power for the plant family, trait complex, functional type, and life form models, respectively. There was a 47% decrease in the number of parameters to estimate for the plant family HOI-inclusive model and 82% for the direct-only model compared to the full HOI-inclusive species identity model. For the trait complex, functional type, and life form models, there was an 85% decrease in the number of parameters to estimate for the HOI-inclusive models and 95% decrease for the direct-only models.

Overall, using the neighbor groupings or abundance models, HOI-inclusive models performed as well as (on the 1:1 line in fig. 2) or better than (below the 1:1 line) the direct-only models for both the Australian and Spain data sets. If using the species identity, rare species grouped, or rare species removed groupings, the direct-only models performed better than the HOI-inclusive models based on  $\text{RMSE}_{\text{test}}$  and BIC and worse than HOI-inclusive models based on  $\text{RMSE}_{\text{train}}$  for both data sets. See tables S3 and S4 for results per focal species for the Australian and Spanish data sets, respectively.

### Discussion

The growing interest in understanding how species interactions within ecological communities modulate coexistence as well as community diversity and stability has driven a reassessment of the assumptions and simplifications commonly made in individual fitness modeling. Clark et al. (2019) found that too much added information can degrade the accuracy of fitness models but also that the best models are not the simplest models commonly used. Though higher-order interactions are widely recognized to be an important group of interactions in natural communities, the issues they cause in modeling individual fitness are not trivial and have driven the opinion that the value of HOIs is not sufficient to merit the issues they cause with modeling (Levine et al. 2017), a conclusion we feel is premature. To improve tractability of HOI-inclusive models, we need to take a hierarchical approach in which, rather than modeling HOIs directly, we first determine which model components are actually important. In this study, we found that focal species identity was most important for intrinsic fecundity. We also found that by grouping neighboring species according

to broad functional categories (specifically, life form and origin status), we can reduce the number of model terms dramatically while improving the fit of HOI-inclusive models. These findings suggest that summarized details about HOIs can be accounted for in individual fitness models without major increases in data demands.

### *Does Identity Matter?*

The identity of individuals interacting in nature is generally considered of central importance to the outcomes of those interactions (Chesson 2000a; Uriarte et al. 2004; Uriarte and Menge 2018). Further, it is assumed that removing species-level information will result in the concurrent removal of associated ecologically important information, resulting in poorly performing models. Though we did find strong evidence that the identity of the focal individual was important in both our Spanish and Australian systems, evidence for the importance of neighbor identity was much weaker. Results from our study suggest that species-specific neighbor identity may be much less important for accurately modeling individual fecundity than previously thought. Notably, we found that even simple information on species characteristics (e.g., life form) maximized model accuracy for the two systems we studied. For most measures of model accuracy ( $\text{RMSE}_{\text{train}}$ ,  $\text{RMSE}_{\text{test}}$ , AIC, and BIC), life form and origin status performed as well as or better than abundance models, which always had the fewest model terms, and the species identity models, which were the most complex. Furthermore, life form and origin status groupings also elevated the performance of HOI-inclusive models relative to direct-only models. These results suggest that it is important to include general ecological information about neighbors (more than just abundance of neighbors), but the finest species-level resolution information is not necessary for capturing ecological differences between interacting neighbors in individual fitness models. This is a noteworthy result because it suggests that the species interactions assumed to be of foundational importance to the maintenance of diversity in plant communities can be summarized for modeling purposes using easy-to-measure functional or life-history groupings (e.g., life form), without the need for precise—and often time-consuming—species identifications. We note that this finding does not speak to the importance of these details for any other question about community-level diversity but does provide a simple guideline for the level of detail needed to maximize the fit of individual fitness models.

The literature on the importance of neighbor identity on focal plant fitness is quite variable. As such, our results align with results from some studies but not others. For instance, Uriarte et al. (2004), Vogt et al. (2010), and

Jacob et al. (2017) found that neighbor identity is not an important driver of interaction outcomes between focal plants and their neighbors in a variety of systems. Other studies, however, have found the opposite, that the identity of the neighbor species does matter, specifically in modeling biomass using pairwise experimental plantings (Weigelt et al. 2002) and when estimating the effect of competition on secondary plant compounds (Barton and Bowers 2006). Our study goes beyond all of these previous neighbor-fitness studies by increasing the number of possible neighbors considered as well as the inclusion of higher-order interactions created by complex neighborhoods. Indeed, the similarities in our neighbor-identity results across two very distinct annual plant systems—semiarid woodlands and ephemeral wetlands—are exciting, as they suggest that some generalizations can be made about how detailed neighborhood information needs to be to accurately estimate individual lifetime fecundity using the full spectrum of relevant plant-plant interactions impacting plants in real natural communities.

### Model Fit

Details of the best-fitting models for each of our two data sets were consistent in some details but not others. Notably, we found that focal identity for intrinsic fecundity was present in the best-performing models for both data sets but not consistently for interaction terms. This result aligns with the well-known fact that there are inherent differences in plant species, likely reflective of the distinct niches held by different species (Tilman 1982; Chesson 2000b; Levine and HilleRisLambers 2009; reviewed in Silvertown 2004). For direct and higher-order interaction model terms, support for the inclusion of focal species identity in models differed for our two systems. The inclusion of focal identity was supported for the interaction terms in the best-fitting model for the Spanish data set but not the Australian data set. In the Australian data set, focal species identity in direct interactions was not important for model performance. This would indicate, at least for the Australian data set, that neighbors had relatively consistent effects on the focal individual, regardless of the focal species. The variation across systems in how important species identity is to fecundity outcomes suggests that the widely held assumption that species differences are important for the outcome of local interactions is actually context dependent.

The importance of HOIs to model fit varied across our data sets. HOI terms were not part of any of the best models for the Australian data set based on  $RMSE_{\text{test}}$ , AIC, or BIC but were for the Spanish data set. HOIs, however, were always included in the best-performing model based on  $RMSE_{\text{train}}$  for both data sets. This indicates that

HOI-inclusive models consistently fit the model to the data better than direct-only models. The exclusion of HOI terms from the best-fitting models for the Australian data set, however, indicates that the inclusion of HOIs into fitness models may not compensate for model complexity based on AIC or BIC for the Australian species. This does not, however, indicate that HOIs are not of significant importance to some focal species, some species pairs, or some species triplets. In this study, we applied HOIs with an all-or-nothing approach, with the assumption that HOIs were either always important or always not important. However, certain HOIs could vary in importance by species in the same way that two species may compete strongly with each other but weakly with a third. For example, intra-specific HOIs could be more important than interspecific HOIs for some species, or vice versa; this would lend itself to a mixed, intermediate model approach that is focal species specific. Such an approach was used in Mayfield and Stouffer (2017b) but was not performed here because we were looking for system-wide trends and were not attempting to provide evidence of the specific importance or value of HOIs to specific species. Our approach is ideal for exploring general principles about fitness modeling but excludes more detailed analysis of the factors influencing specific species and specific interactions. Indeed, Mayfield and Stouffer (2017b), which used data on the same Australian focal species we use here, examined the impacts of HOIs on each focal species separately and found that HOIs were significantly important for half of our tested focal species but not all, supporting our conclusion about the variable importance of HOIs in this Australian data set. In the Spanish data set, two high-performing models for the focal identity analysis were identified: full model (direct) and full model (direct and HOI). Full model (direct) was the best-fitting model based on  $RMSE_{\text{test}}$  and BIC (a fit-focused metric that heavily penalizes for number of model terms), while full model (direct and HOI) was the top-performing model based on  $RMSE_{\text{train}}$  and AIC. The similar performance of these models, one including HOIs and one excluding them, may suggest that HOIs are less important for some focal species in this system than others and that the overall HOI signal may have become lost when balanced against the number of parameters used to predict fitness using some tests of this entire data set.

Though our general conclusion is that HOIs are of variable importance to species in both of our focal systems, it is interesting that evidence for this manifests in such different ways for the two systems. We did select annual plant systems that differed in substantial ways to increase the value of our results in making generalizable conclusions. Notably, our systems are from opposite sides of the globe and thus have totally distinct evolutionary and biogeographic histories. The Australian system is a



semiarid system in which water is limiting and soil phosphorous is a major factor determining diversity patterns (Prober and Wiehl 2011; Dwyer et al. 2015). This system has also been heavily impacted by human activities, including the invasion of grass and forb species from Africa, Europe, and North America, as well as fragmentation, soil nutrient enrichment, and disturbance (Prober and Wiehl 2011; Dwyer et al. 2015). Our Spanish system, on the other hand, is an ephemeral wetland system, with diversity structured in relation to inundation level and extent (Lanuza et al. 2018). It is a fully native system that sits in a well-preserved nature reserve protected from many of the anthropogenic factors impacting the Australian system. Given the major differences in these systems, it is perhaps not surprising that results differed between the systems. In fact, given these differences, it is possible that our results reflect important inherent differences in these systems. The Australian data set has more than twice the number of potential neighbors as the Spanish data set, resulting in a corresponding doubling of the number of possible HOI terms. The Australian plant community was also 2.2 times as even (based on Pielou's evenness index; Pielou 1966) than the Spanish data set, which was dominated by a single species, *Hordeum marinum* (with more than 10 times as many individuals of this species than any other; table S5). The dominance of *H. marinum* strongly influenced which direct interactions and HOIs could be parameterized in the Spanish models. (We note again that in all models run in this study, only those interactions for which data were available were parameterized; models did not include all possible interaction terms as not all combinations occurred in our system). It is important to never lose sight of the fact that all HOI-inclusive individual fitness models (at least using this approach to their inclusion) are simplified versions of the theoretical ideal of a full HOI-inclusive model. This is because no data set includes data on the full suite of all possible direct and HOIs for all species in a natural community. Such a data set is only even possible for data collected from a fully artificial experiment; in natural systems, many species found in the same community may never interact (at least as adults) in local neighborhoods. Even those models with several hundred terms have parameters for a fraction of all possible HOIs. Our results again point to the need to consider what information about HOIs is most important to include for the goals of a specific study or question.

#### *Functional versus Species Identity*

The main aim of this study was to determine whether HOI-inclusive models could be simplified in a way that reduced vulnerability to model overfitting while retaining

any value added by including this biological complexity into individual fitness models. Thus, even though we did not find HOIs to be essential components of all of the very best-fitting models for both systems, it is still worth exploring how well HOI-inclusive models performed when simplified in a variety of ecologically sensible ways. We found that HOI-inclusive models in both systems that used trait-based or functional group categories of neighboring species were more supported than those including all parameterizable species pairs. Further, certain function-based neighbor groupings performed better than the neighbor abundance models, which assumed all neighbors were the same species, which had the fewest model terms. These results are consistent with much of the functional trait literature, which has made strong links between traits, performance, and/or demographic rates (Westoby and Wright 2006; Poorter et al. 2008; Violle et al. 2017). The relationship between functional trait trade-offs and coexistence is well studied (Ehrlén et al. 1998; Kneitel and Chase 2004; Angert et al. 2009; Sterck et al. 2011; Soliveres et al. 2014; Kraft et al. 2015; Petry et al. 2018), and many functional traits are known to affect how species interact (Bennett et al. 2016; Kunstler et al. 2016; Rolhauser and Pucheta 2016; Pérez-Ramos et al. 2019). Given the depth of literature linking function to demographic success, our results are not surprising. To our knowledge, however, our study is the first explicit test of whether species that are broadly functionally similar also contribute to higher-order interactions in similar ways. As is the case for the link between function and direct interactions (Angert et al. 2009; Kraft et al. 2015; Kunstler et al. 2016), we might expect that grouping species by function would be effective only if we could identify the traits specifically involved in the mechanisms driving HOIs. More detailed study of specific mechanisms underlying HOIs will likely show that grouping neighbors by specific HOI mechanism or associated traits is the best option, but our results show that even broad general life-history strategies are useful for grouping neighboring species in HOI-inclusive fitness models. From a practical perspective, this finding is exciting because it shows that the common reason given for excluding HOIs from fitness models—that data requirements are too great—is overstated (Levine et al. 2017; AlAdwani and Saavedra 2019). Basic information about the abundance of neighbors and their life history (grass/forb) or basic functional type is sufficient to gain the benefits of adding HOIs to individual fitness models when they prove strong in a given system. For example, if using the life form grouping in the systems studied here, neighboring plant surveys could be simplified to counts of grass or forb plants and not specific species' identities. In addition, with the reduced number of species-specific terms required to sustain the power of these models, researchers could include other

types of information not included here, such as environmental variability and multitrophic interactions without the model becoming unwieldy (Bimler et al. 2018; Godoy et al. 2018). Given that we admittedly explored only gross functional categories based on data that we had available for two distinct data sets, we suggest that our results are the tip of the iceberg, and it would be very interesting to conduct more detailed studies of the relationships between functional traits, HOI mechanisms, and the overall importance of HOIs to individual fecundity outcomes.

### *The Role of Rare Species*

Rare species are a common feature of all communities (MacArthur and Wilson 1967). Because they are, as a group, ubiquitous, it is valuable to know whether detailed information about rare species within communities is important for accurately estimating or predicting fitness outcomes for individual plants living in natural communities. Rare species are often absent from data sets because they are rare and often missed in nonexhaustive surveys or are systematically excluded due to difficulties involved in identifying and cataloging what can be a large number of taxa that only appear once or twice. Our results suggest that excluding or grouping rare species together when they appear in neighborhoods is a viable solution for improving model tractability without seriously reducing the fit or predictive power of individual fitness models. More research, however, is needed to fully understand the importance of rare species to the fitness of individual species in specific systems.

### *Limitations*

In this study, we focus on biologically sensible ways of balancing realistic data requirements with more complex models that incorporate more biological realism than standard fitness models. We suggest exploring functional groupings as a way to do this. Grouping species into broader functional categories does, however, eliminate some intra- and interspecific variability that may be of great interest and importance when modeling coexistence or other aspects of diversity maintenance (Jung et al. 2010; Violle et al. 2012; Hart et al. 2016; Uriarte and Menge 2018). As all models are simplifications of reality, the question remains: how important is each piece of information in a model and how much is lost when specific pieces of information are excluded? Given the comparatively strong performance of HOI-inclusive models in this study, it becomes important for future research to consider what questions the fitness model outcomes are being used to answer and whether trade-offs in accuracy and simplicity are acceptable. In this study, we aimed to determine

whether there were biologically sensible approaches to reducing the number of terms in HOI-inclusive fitness models without losing the modeling benefits of the inclusion of HOIs in the first place. Our study outcomes align with other studies showing that HOIs are important for seed set-based fitness outcomes for some species (Mayfield and Stouffer 2017b). We also found, however, that across quite different plant communities, HOI-inclusive model power can be retained by grouping neighboring species by life form or broad functional type. That said, much species-specific information is certainly lost through this simplification process, which may be important if the end goal of estimating fitness is to determine something detailed about a specific species such as coexistence potential. In such cases, it would likely be beneficial to leave details about target species pairs (focal and neighbor) in the model as separate terms, while grouping all other neighbors by life form or function, as supported by our findings. This approach is similar to that used for coexistence modeling using direct-interaction models (Levine and HilleRisLambers 2009; Wainwright et al. 2019).

If sufficient data are available to include species-level information in HOI-inclusive models without model overfitting, the use of more information can absolutely add important details, such as when HOIs drive changes in the magnitude and direction of direct, pairwise interactions. The inclusions of more HOI terms may have this effect through accounting for the impacts of second-order neighbor effects. For instance, by comparing the estimated coefficients of our abundance and life form models for both data sets (plus origin status and trait complex for the Australian data set), we see that the direct interaction coefficient values change in magnitude and sometimes sign when HOIs are included (app. S4). Some of this would also be evident with more coarsely grouped HOI terms, but obviously incremental losses in information will accompany all model simplifications. Trade-offs between ecological detail and poor model performance and overfitting will always exist, but we hope this study will help guide researchers to make justified decisions about how to simplify HOI terms when modeling individual fitness.

Further, we note that this work does not increase our understanding of the mechanisms underlying HOIs in these communities. We could speculate that because functional identity, including life form (i.e., grass vs. forb), was important to both systems, competition for water in these Mediterranean systems might be an underlying mechanism of HOIs related to differences in root morphology (Ravenek et al. 2016) and leaf physiology (Pérez-Ramos et al. 2019). However, explicit experiments are needed to test this hypotheses and other underlying mechanisms of HOIs. Similar analyses to the one used here can be used on other measures of plant performance, such as biomass

or basal area. This would help to expand our understanding of HOIs and driving mechanisms in perennial plant communities. By further exploration of the above, we could continue work toward identifying generalizable trends and mechanisms across plant communities of differing compositions.

### Conclusions

In order to balance realistic data requirements with complex models, we need meaningful approaches to direct model simplification. Here, we have identified defensible approaches to tackle this complexity by including or removing focal identity in models as well as grouping neighbors by function rather than species identity. We presented a series of simplifications in model complexity that prove useful to understand the effect of HOIs on individual fitness. These simplifications can improve the accuracy of coexistence and diversity modeling while maintaining feasibly sized field studies.

### Acknowledgments

M.M.M. is grateful for support provided by the Australian Research Council (DP170100837). D.B.S. acknowledges the support of a Marsden Grant and a Rutherford Discovery Fellowship (16-UOC-008 and RDF-13-UOC-003), administered by the Royal Society of New Zealand Te Apārangi. O.G. acknowledges support provided by the Ramón y Cajal Programme (RYC-2017-23666). I.B. and O.G. also acknowledge financial support provided by the Spanish Ministry of Economy and Competitiveness Explora Program (CGL2017-92436-EXP, SIMPLEX) and Retos Investigación (RTI2018-098888-A-I00, MeDiNaS). We thank Doñana National Park staff members for granting access to Caracoles real estate and Xingwen Loy for use of his plant designs in figure 1.

### Statement of Authorship

Ideas in this study were conceived and developed by M.M.M., D.B.S., and T.E.M. Data collection was performed by M.M.M., O.G., and I.B. Code was developed by D.B.S. and modified by T.E.M. for the particular analyses. Models were run and analyzed by T.E.M. The manuscript was written by T.E.M. under the guidance of M.M.M., with significant contributions from and edits by all authors. All authors read and approved the final manuscript.

### Data and Code Availability

Trait and fecundity data for both the Australia and Spanish data sets have been deposited in the Dryad Digital Reposi-

tory (<https://doi.org/10.5061/dryad.zs7h44j7f>; Martyn et al. 2020). Code for analyses and figure generation are available at [https://github.com/tmartyn/Balance\\_Accuracy\\_Similarity](https://github.com/tmartyn/Balance_Accuracy_Similarity) and <https://doi.org/10.5281/zenodo.43216>.

### Literature Cited

- Abrams, P. A. 1983. Arguments in favor of higher-order interactions. *American Naturalist* 121:887–891.
- Adler, P. B., J. HilleRisLambers, and J. M. Levine. 2007. A niche for neutrality. *Ecology Letters* 10:95–104.
- Adler, P. B., D. Smull, K. H. Beard, R. T. Choi, T. Furniss, A. Kulmatiski, J. M. Meiners, A. T. Tredennick, and K. E. Veblen. 2018. Competition and coexistence in plant communities: intra-specific competition is stronger than interspecific competition. *Ecology Letters* 21:1319–1329.
- AlAdwani, M., and S. Saavedra. 2019. Is the addition of higher-order interactions in ecological models increasing the understanding of ecological dynamics? *Mathematical Biosciences* 315:108222.
- Angert, A. L., T. E. Huxman, P. Chesson, and D. L. Venable. 2009. Functional tradeoffs determine species coexistence via the storage effect. *Proceedings of the National Academy of Sciences of the USA* 106:11641–11645.
- Bairey, E., E. D. Kelsic, and R. Kishony. 2016. High-order species interactions shape ecosystem diversity. *Nature Communications* 7:1–7.
- Barton, K. E., and M. D. Bowers. 2006. Neighbor species differentially alter resistance phenotypes in *Plantago*. *Oecologia* 150:442–452.
- Bennett, J. A., K. Riibak, R. Tamme, R. J. Lewis, and M. Pärtel. 2016. The reciprocal relationship between competition and intraspecific trait variation. *Journal of Ecology* 104:1410–1420.
- Billick, I., and T. J. Case. 1994. Higher order interactions in ecological communities: what are they and how can they be detected? *Ecology* 75:1529–1543.
- Bimler, M. D., D. B. Stouffer, H. R. Lai, and M. M. Mayfield. 2018. Accurate predictions of coexistence in natural systems require the inclusion of facilitative interactions and environmental dependency. *Journal of Ecology* 106:1839–1852.
- Brewer, M. J., A. Butler, and S. L. Cooksley. 2016. The relative performance of AIC, AIC<sub>c</sub> and BIC in the presence of unobserved heterogeneity. *Methods in Ecology and Evolution* 7:679–692.
- Bulleri, F., J. F. Bruno, B. R. Silliman, and J. J. Stachowicz. 2016. Facilitation and the niche: implications for coexistence, range shifts and ecosystem functioning. *Functional Ecology* 30:70–78.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference. 2nd ed. Springer, New York.
- Callaway, R. M., B. E. Mahall, C. Wicks, J. Pankey, and C. Zabinski. 2003. Soil fungi and the effects of an invasive forb on grasses: neighbor identity matters. *Ecology* 84:129–135.
- Case, T. J., and E. A. Bender. 1981. Testing for higher order interactions. *American Naturalist* 118:920–929.
- Cavender-Bares, J., K. H. Kozak, P. V. A. A. Fine, and S. W. Kembel. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters* 12:693–715.
- Chesson, P. 2000a. General theory of competitive coexistence in spatially varying environments. *Theoretical Population Biology* 58:211–237.
- . 2000b. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics* 31:343–366.

- Clark, A. T., L. Ann Turnbull, A. Tredennick, E. Allan, W. S. Harpole, M. M. Mayfield, S. Soliveres, et al. 2019. Predicting species abundances in a grassland biodiversity experiment: trade-offs between model complexity and generality. *Journal of Ecology*, <https://doi.org/10.1111/1365-2745.13316>.
- Craine, J. M., J. Froehle, D. G. Tilman, D. A. Wedin, and F. S. Chapin. 2001. The relationships among root and leaf traits of 76 grassland species and relative abundance along fertility and disturbance gradients. *Oikos* 93:274–285.
- Díaz, S., J. Kattge, J. H. C. Cornelissen, I. J. Wright, S. Lavorel, S. Dray, B. Reu, et al. 2016. The global spectrum of plant form and function. *Nature* 529:167–171.
- Dwyer, J. M., R. J. Hobbs, C. E. Wainwright, and M. M. Mayfield. 2015. Climate moderates release from nutrient limitation in natural annual plant communities. *Global Ecology and Biogeography* 24:549–561.
- Ehrlén, J., J. M. van Groenendaal, and V. Groenendaal. 1998. The trade-off between dispersability and longevity—an important aspect of plant species diversity. *Applied Vegetation Science* 1:29–36.
- Engel, E. C., and J. F. Weltzin. 2008. Can community composition be predicted from pairwise species interactions? *Plant Ecology* 195:77–85.
- Eskelinen, A. 2008. Herbivore and neighbour effects on tundra plants depend on species identity, nutrient availability and local environmental conditions. *Journal of Ecology* 96:155–165.
- Fowler, N. 1981. Competition and coexistence in a North Carolina grassland. II. The effects of the experimental removal of species. *Journal of Ecology* 69:843–854.
- Gallien, L., N. E. Zimmermann, J. M. Levine, and P. B. Adler. 2017. The effects of intransitive competition on coexistence. *Ecology Letters* 20:791–800.
- García-Callejas, D., O. Godoy, and I. Bartomeus. 2020. *cxr*: a toolbox for modelling species coexistence in R. *Methods in Ecology and Evolution*, <https://doi.org/10.1111/2041-210X.13443>.
- Gellner, G., and K. S. McCann. 2016. Consistent role of weak and strong interactions in high- and low-diversity trophic food webs. *Nature Communications* 7:1–7.
- Godoy, O., I. Bartomeus, R. P. Rohr, and S. Saavedra. 2018. Towards the integration of niche and network theories. *Trends in Ecology and Evolution* 33:287–300.
- Godoy, O., N. J. B. B. Kraft, and J. M. Levine. 2014. Phylogenetic relatedness and the determinants of competitive outcomes. *Ecology Letters* 17:836–844.
- Godoy, O., D. B. Stouffer, N. J. B. B. Kraft, and J. M. Levine. 2017. Intransitivity is infrequent and fails to promote annual plant coexistence without pairwise niche differences. *Ecology* 98:1193–1200.
- Goldberg, D. E., and P. A. Werner. 1983. Equivalence of competitors in plant communities: a null hypothesis and a field experimental approach. *American Journal of Botany* 70:1098–1104.
- Hart, S. P., S. J. Schreiber, and J. M. Levine. 2016. How variation between individuals affects species coexistence. *Ecology Letters* 19:825–838.
- Houlahan, J. E., S. T. McKinney, T. M. Anderson, and B. J. McGill. 2017. The priority of prediction in ecological understanding. *Oikos* 126:1–7.
- Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ.
- Jacob, C. E., E. Tozzi, and C. J. Willenborg. 2017. Neighbour presence, not identity, influences root and shoot allocation in pea. *PLoS ONE* 12:1–12.
- Jung, V., C. Violle, C. Mondy, L. Hoffmann, and S. Muller. 2010. Intraspecific variability and trait-based community assembly. *Journal of Ecology* 98:1134–1140.
- Kaitaniemi, P., and A. Lintunen. 2010. Neighbor identity and competition influence tree growth in Scots pine, Siberian larch, and silver birch. *Annals of Forest Science* 67:1–9.
- Kimball, S., J. L. Funk, M. J. Spasojevic, K. N. Suding, S. Parker, and M. L. Goulden. 2016. Can functional traits predict plant community response to global change? *Ecosphere* 7:1–18.
- Kleinheselink, A. R., N. J. Kraft, and J. M. Levine. 2020. Mechanisms underlying higher order interactions: from quantitative definitions to ecological processes. *bioRxiv*, <https://doi.org/10.1101/857920>.
- Kneitel, J. M., and J. M. Chase. 2004. Trade-offs in community ecology: linking spatial scales and species coexistence. *Ecology Letters* 7:69–80.
- Kraft, N. J. B., O. Godoy, and J. M. Levine. 2015. Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences of the USA* 112:797–802.
- Kuhn, M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28:1–26.
- Kunstler, G., D. Falster, D. A. Coomes, F. Hui, R. M. Kooyman, D. C. Laughlin, L. Poorter, et al. 2016. Plant functional traits have globally consistent effects on competition. *Nature* 529:204–207.
- Lanuza, J. B., I. Bartomeus, and O. Godoy. 2018. Opposing effects of floral visitors and soil conditions on the determinants of competitive outcomes maintain species diversity in heterogeneous landscapes. *Ecology Letters* 21:865–874.
- Letten, A. D., and D. B. Stouffer. 2019. The mechanistic basis for higher-order interactions and non-additivity in competitive communities. *Ecology Letters* 22:423–436.
- Levine, J. M. 2016. Ecology: a trail map for trait-based studies. *Nature* 529:163–164.
- Levine, J. M., J. Bascompte, P. B. Adler, and S. Allesina. 2017. Beyond pairwise mechanisms of species coexistence in complex communities. *Nature* 546:56–64.
- Levine, J. M., and J. HilleRisLambers. 2009. The importance of niches for the maintenance of species diversity. *Nature* 461:254–257.
- Levine, J. M., M. Vilà, C. M. D. Antonio, J. S. Dukes, K. Grigulis, and S. Lavorel. 2003. Mechanisms underlying the impacts of exotic plant invasions. *Proceedings of Royal Society B* 270:775–781.
- Li, Y., M. M. Mayfield, B. Wang, J. Xiao, K. Kral, D. Janik, J. Holik, and C. Chu. 2020. Beyond direct neighbourhood effects: higher-order interactions improve modelling and predicting tree survival and growth. *National Science Review*, <https://doi.org/10.1093/nsr/nwaa244>.
- Li, Y., B. A. Schichtel, J. T. Walker, D. B. Schwede, X. Chen, C. M. B. Lehmann, M. A. Puchalski, D. A. Gay, and J. L. Collett. 2016. Increasing importance of deposition of reduced nitrogen in the United States. *Proceedings of the National Academy of Sciences of the USA* 113:5874–5879.
- Lübke, T., B. Schuldt, and C. Leuschner. 2015. Species identity and neighbor size surpass the impact of tree species diversity on productivity in experimental broad-leaved tree sapling assemblages under dry and moist conditions. *Frontiers in Plant Science* 6:1–12.



- MacArthur, R. H. 1972. *Geographical ecology: patterns in the distribution of species*. Harper and Row, New York.
- MacArthur, R. H., and E. O. Wilson. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, NJ.
- Mack, R. N., and J. L. Harper. 1977. Interference in dune annuals: spatial pattern and neighbourhood. *Journal of Ecology* 65:345–363.
- Maris, V., P. Huneman, A. Coreau, S. Kéfi, R. Pradel, and V. Devictor. 2018. Prediction in ecology: promises, obstacles and clarifications. *Oikos* 127:171–183.
- Martorell, C., and R. P. Freckleton. 2014. Testing the roles of competition, facilitation and stochasticity on community structure in a species-rich assemblage. *Journal of Ecology* 102:74–85.
- Martyn, T., D. Stouffer, O. Godoy, I. Bartomeus, A. Pastore, and M. Mayfield. 2020. Data from: Identifying “useful” fitness models: balancing the benefits of added complexity with realistic data requirements in models of individual plant fitness. *American Naturalist*, Dryad Digital Repository, <https://doi.org/10.5061/dryad.zs7h44j7f>.
- Matías, L., O. Godoy, L. Gómez-Aparicio, and I. M. Pérez-Ramos. 2018. An experimental extreme drought reduces the likelihood of species to coexist despite increasing intransitivity in competitive networks. *Journal of Ecology* 106:826–837.
- Mayfield, M. M., and D. B. Stouffer. 2017a. Data from: Higher-order interactions capture unexplained complexity in diverse communities. *American Naturalist*, Dryad Digital Repository, <https://doi.org/10.5061/dryad.3562g>.
- . 2017b. Higher-order interactions capture unexplained complexity in diverse communities. *Nature Ecology and Evolution* 1:1–7.
- McCann, K., A. Hastings, and G. R. Huxel. 1998. Weak trophic interactions and the balance of nature. *Nature* 395:794–798.
- McCullagh, P., and J. A. Nelder. 1989. *Generalized linear models*. Monographs on statistics and applied probability. Vol 37. 2nd ed. Chapman and Hall, London.
- McGill, B. J., B. J. Enquist, E. Weiher, and M. Westoby. 2006. Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution* 21:178–185.
- Michalet, R., J. P. Maalouf, P. Choler, B. Clément, D. Rosebery, J. M. Royer, C. Schöb, and C. J. Lortie. 2014. Competition, facilitation and environmental severity shape the relationship between local and regional species richness in plant communities. *Ecography* 38:335–345.
- Moen, J. 1989. Diffuse competition: a diffuse concept. *Oikos* 54: 260–263.
- Morin, P. J., S. P. Lawler, and E. A. Johnson. 1988. Competition between aquatic insects and vertebrates: interaction strength and higher order interactions. *Ecology* 69:1401–1409.
- Neill, W. E. 1974. The community matrix and interdependence of the competition coefficients. *American Naturalist* 108:399–408.
- Ordóñez, A., I. J. Wright, and H. Olff. 2010. Functional differences between native and alien species: a global-scale comparison. *Functional Ecology* 24:1353–1361.
- Paterson, R. A., J. T. A. Dick, D. W. Pritchard, M. Ennis, M. J. Hatcher, and A. M. Dunn. 2015. Predicting invasive species impacts: a community module functional response approach reveals context dependencies. *Journal of Animal Ecology* 84:453–463.
- Pérez-Ramos, I. M., L. Matías, L. Gómez-Aparicio, and Á. Godoy. 2019. Functional traits and phenotypic plasticity modulate species coexistence across contrasting climatic conditions. *Nature Communications* 10:1–11.
- Petry, W. K., G. S. Kandlikar, N. J. Kraft, O. Godoy, and J. M. Levine. 2018. A competition-defence trade-off both promotes and weakens coexistence in an annual plant community. *Journal of Ecology* 106:1806–1818.
- Pielou, E. C. 1966. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology* 13:131–144.
- Pomerantz, M. J. 1981. Do “higher order interactions” in competition systems really exist? *American Naturalist* 117:583–591.
- Poorter, L., S. J. Wright, H. Paz, D. D. Ackerly, R. Condit, G. Ibarra-Manríquez, K. E. Harms, et al. 2008. Are functional traits good predictors of demographic rates? evidence from five Neotropical forests. *Ecology* 89:1908–1920.
- Prober, S. M., and G. Wiehl. 2011. Resource heterogeneity and persistence of exotic annuals in long-ungrazed Mediterranean-climate woodlands. *Biological Invasions* 13:2009–2022.
- Qian, H., and L. Jiang. 2014. Phylogenetic community ecology: integrating community ecology and evolutionary biology. *Journal of Plant Ecology* 7:97–100.
- R Core Development Team. 2018. R: a language and environment for statistical computing: 3.5.1. R Foundation for Statistical Computing, Vienna.
- Ravenek, J. M., L. Mommer, E. J. W. Visser, J. van Ruijven, J. W. van der Paauw, A. Smit-Tiekstra, H. de Caluwe, and H. de Kroon. 2016. Linking root traits and competitive success in grassland species. *Plant and Soil* 407:39–53.
- Rollhauser, A. G., and E. Pucheta. 2016. Annual plant functional traits explain shrub facilitation in a desert community. *Journal of Vegetation Science* 27:60–68.
- Silvertown, J. 2004. Plant coexistence and the niche. *Trends in Ecology and Evolution* 19:605–611.
- Soliveres, S., F. T. Maestre, M. A. Bowker, R. Torices, J. L. Quero, M. García-Gómez, O. Cabrera, et al. 2014. Functional traits determine plant co-occurrence more than environment or evolutionary relatedness in global drylands. *Perspectives in Plant Ecology, Evolution and Systematics* 16:164–173.
- Soliveres, S., F. T. Maestre, W. Ulrich, P. Manning, S. Boch, M. A. Bowker, D. Prati, et al. 2015. Intransitive competition is widespread in plant communities and maintains their species richness. *Ecology Letters* 18:790–798.
- Sterck, F., L. Markesteijn, F. Schieving, and L. Poorter. 2011. Functional traits determine trade-offs and niches in a tropical forest community. *Proceedings of the National Academy of Sciences of the USA* 108:20627–20632.
- Straub, C. S., and W. E. Snyder. 2006. Species identity dominates the relationship between predator biodiversity and herbivore suppression. *Ecology* 87:277–282.
- Tilman, D. G. 1982. *Resource competition and community structure*. Princeton University Press, Princeton, NJ.
- Tjoelker, M. G., J. M. Craine, D. Wedin, P. B. Reich, and D. Tilman. 2005. Linking leaf and root trait syndromes among 39 grassland and savannah species. *New Phytologist* 167:493–508.
- Uriarte, M., R. Condit, C. D. Canham, and S. P. Hubbell. 2004. A spatially explicit model of sapling growth in a tropical forest: does the identity of neighbours matter? *Journal of Ecology* 92:348–360.
- Uriarte, M., and D. Menge. 2018. Variation between individuals fosters regional species coexistence. *Ecology Letters* 21:1496–1504.



- Vandermeer, J. H., and N. May. 1969. The competitive structure of communities: an experimental approach with protozoa. *Ecology* 50:362–371.
- Vázquez, D. P., C. J. Melián, N. M. Williams, N. Blüthgen, B. R. Krasnov, and R. Poulin. 2007. Species abundance and asymmetric interaction strength in ecological networks. *Oikos* 116:1120–1127.
- Violle, C., B. J. Enquist, B. J. McGill, L. Jiang, C. H. Albert, C. Hulshof, V. Jung, and J. Messier. 2012. The return of the variance: intraspecific variability in community ecology. *Trends in Ecology and Evolution* 27:244–252.
- Violle, C., W. Thuiller, N. Mouquet, F. Munoz, N. J. B. Kraft, M. W. Cadotte, S. W. Livingstone, and D. Mouillot. 2017. Functional rarity: the ecology of outliers. *Trends in Ecology and Evolution* 32:356–367.
- Vogt, D. R., D. J. Murrell, and P. Stoll. 2010. Testing spatial theories of plant coexistence: no consistent differences in intra- and interspecific interaction distances. *American Naturalist* 175:73–84.
- Volterra, V. 1928. Variations and fluctuations of the number of individuals in animal species living together. *ICES Journal of Marine Science* 3:3–51.
- Wainwright, C. E., J. HilleRisLambers, H. R. Lai, X. Loy, and M. M. Mayfield. 2019. Distinct responses of niche and fitness differences to water availability underlie variable coexistence outcomes in semi-arid annual plant communities. *Journal of Ecology* 107:293–306.
- Wang, Y., U. Naumann, S. T. Wright, and D. I. Warton. 2012. Mvabund—an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution* 3:471–474.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33:475–505.
- Weigelt, A., T. Steinlein, and W. Beyschlag. 2002. Does plant competition intensity rather depend on biomass or on species identity? *Basic and Applied Ecology* 3:85–94.
- Westoby, M., and I. J. Wright. 2006. Land-plant ecology on the basis of functional traits. *Trends in Ecology and Evolution* 21:261–268.
- White, E. P., P. B. Adler, W. K. Lauenroth, R. A. Gill, D. Greenberg, D. M. Kaufman, A. Rassweiler, et al. 2006. A comparison of the species-time relationship across ecosystems and taxonomic groups. *Oikos* 112:185–195.
- Wilbur, H. M. 1972. Competition, predation, and the structure of the *Ambystoma–Rana sylvatica* community. *Ecology* 53:3–21.
- Wilcox, T. M., M. K. Schwartz, and W. H. Lowe. 2018. Evolutionary community ecology: time to think outside the (taxonomic) box. *Trends in Ecology and Evolution* 33:240–250.
- Wissinger, S., and J. Mcgrady. 1993. Intraguild predation and competition between larval dragonflies: direct and indirect effects on shared prey. *Ecology* 74:207–218.
- Wootton, J. T. 1994. The nature and consequences of indirect effects in ecological communities. *Annual Review of Ecology and Systematics* 25:443–466.
- You, C., F. Wu, Y. Gan, W. Yang, Z. Hu, Z. Xu, B. Tan, L. Liu, and X. Ni. 2017. Grass and forbs respond differently to nitrogen addition: a meta-analysis of global grassland ecosystems. *Scientific Reports* 7:1–10.
- Zhang, P. 1993. Model selection via multifold cross validation. *Annals of Statistics* 21:299–313.

### References Cited Only in the Online Enhancements

- Cornelissen, J. H., S. Lavorel, E. Garnier, S. Díaz, N. Buchmann, D. E. Gurvich, P. B. Reich, et al. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian Journal of Botany* 51:335–380.
- Dwyer, J. M., R. J. Hobbs, and M. M. Mayfield. 2014. Specific leaf area responses to environmental gradients through space and time. *Ecology* 95:399–410.
- Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik. 2018. cluster: cluster analysis basics and extensions. <https://CRAN.R-project.org/package=cluster>.
- Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53–65.
- Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423.
- Van der Laan, M. J., K. S. Pollard, and J. Bryan. 2003. A new partitioning around medoids algorithm. *Journal of Statistical Computation and Simulation* 73:575–584.

Associate Editor: David Vasseur  
Editor: Daniel I. Bolnick